Structural insight into co-translational membrane protein folding

Grant A. Pellowe, Paula J. Booth*

Department of Chemistry, King's College London, 7 Trinity Street, SE1 1DB, London, UK
* Corresponding author: paula.booth@kcl.ac.uk

Abstract

Membrane protein folding studies lag behind those of water-soluble proteins due to immense difficulties of experimental study, resulting from the need to provide a hydrophobic lipid-bilayer environment when investigated *in vitro*. A sound understanding of folding mechanisms is important for membrane proteins as they contribute to a third of the proteome and are frequently associated with disease when mutated and/or misfolded. Membrane proteins largely consist of α-helical, hydrophobic transmembrane domains, which insert into the membrane, often using the SecYEG/Sec61 translocase system. This mini-review highlights recent advances in techniques that can further our understanding of co-translational folding and notably, the structure and insertion of nascent chains as they emerge from translating ribosomes.

Abbreviations

SRP: signal recognition particle, RNC: ribosome-bound nascent chain complex, HTL: holotranslocon, ATP: adenosine-triphosphate, Hsp: heat-shock protein, TRAM: translocon-associated membrane protein, Trap: translocon-associated protein, OST: oligosaccaryltransferase, FRET: Förster resonance energy transfer, TM: transmembrane, MP: membrane protein, PURE: protein synthesis using recombinant elements, NC: nascent chain, PTC: peptidyl transferase centre, CMC: critical micelle concentration, IMV: inner membrane vesicle, MSP: membrane scaffold protein, SMA: styrene/maleic acid, DIBMA: diisobutylene/maleic acid, UAA: unnatural amino acid, AMS: 4-Acetamido-4'maleimidylstilbene-2,2'-disulfonic acid disodium salt, AP: arrest-peptide, SEIRAS: surface-enhanced infrared spectroscopy, RRL: rabbit reticulocyte lysate, PE: phosphatidylethanolamine, ER: endoplasmic reticulum, PMF: proton motive force, DMPC: 1,2-dimyristoyl-sn-glycero-3-phosphocholine

1. Membrane protein folding in vivo

The majority of *Escherichia coli* α-helical nascent chains follow the Sec-dependent pathway of insertion, and are captured by the signal recognition particle (SRP) and delivered towards its receptor FtsY for the binding of the ribosome nascent *chain* (RNC) to the SecYEG insertion apparatus [1]. SecYEG with associated chaperones is often termed the holotranslocon (HTL); this larger complex is composed of SecYEG-SecDFYajC-YidC [2], and acts as a protein channel, translocase and insertase and protects the nascent chain from aggregation as it begins to fold on the cytoplasmic side of the membrane [3, 4]. In eukaryotic organisms, a hetero-trimeric complex Sec61, of three subunits (α, β, γ) is essential for both co-translational insertion of membrane proteins and translocation of secretory proteins into, and across the endoplasmic reticulum membrane for cellular trafficking. Sec62 and Sec63 bind during a post-translational event in yeast, and most likely other eukaryotes, to ratchet peptides through the Sec61 channel with the aid of BiP, an ATPase of the Hsp70 family [5]. This mechanism is similar to protein secretion mechanisms across the bacterial inner membrane, however SecA pushes the peptide through the channel in an ATP-dependent manner [1]. Sec61 also associates itself with chaperoning proteins in higher-order complexes like translocon-associated membrane protein (TRAM), translocon-associated protein (Trap) and oligosaccharyltransferase (OST), which together are required for the folding of glycosylated proteins in eukaryotes [5].

TMs of polytopic membrane proteins are generally assumed to insert into the membrane sequentially, however more complex scenarios suggest that insertion can also occur as two-helix hairpins which are

assembled close to the membrane or even in the vicinity of the Sec61/SecY channels before their coordinated release into the membrane [6-9]. The helical structure can form in the ribosome exit tunnel, as shown by Förster resonance energy transfer (FRET), gel shift assays and structural studies [10]. A long standing model of translocon function proposes that transmembrane TM helices insert into the bilayer via a channel and lateral gate in the translocon [11]. This model is supported by chemical-crosslinking studies whereby stalled nascent chains were found to crosslink to both the translocon and surrounding lipids, in isolated ER-microsomes [12]. A recent hypothesis suggests that TM helices do not necessarily enter the SecYEG channel, but slide down the outside of the translocon, inserting via lipid head groups before the hydrophobic interior of the membrane provides the driving for insertion of the TMs [11]. More polar regions of the protein may use the translocon channel to insert into, or cross the bilayer. The chemistry of the lipid head groups, packing of hydrophobic tails and overall lipid bilayer properties are important for the insertion of nascent TM helices in either model.

Smaller and less complex proteins have been observed to insert with only YidC-type insertases in a 'Sec-independent' manner. Substrates for YidC, or Oxa1 of mitochondria or Alb3/Alb4 of chloroplasts are likely to favour smaller proteins with one or two helices [13]. The precise function of the YidC-like proteins is currently unknown, with only a direct catalytic activity observed thus far when reconstituted *in vitro* [14]. Mitochondria do not contain the Sec machinery and only contain Oxa1, suggesting that direct contact with lipids is essential for MP insertion and spontaneous folding can occur with the insertase partitioning charged/highly polar residues across the membrane [11]. However, Oxa1 may cooperate with the TIM23 machinery and also Mba1 during membrane protein insertion, forming a larger insertion complex like HTL in E. coli, therefore it remains unclear whether YidC/Oxa1 act as standalone insertases, or remain as a small unknown part of a whole [15].

2. Tools required for co-translational study

2.1. Synthesis machinery

Cell-free translation systems **(figure 1a)** are often used to study co-translational folding due to the ease of manipulation of translation conditions and synthesis components, such as amino acids. Cell-free systems were originally derived from cell lysates of a specific expression organism. *E. coli* strains optimised for protein expression had their lysate harvested giving rise to the S30 expression system. Endogenous membranes have been removed from these extracts during their preparation, therefore for the expression of MPs a membrane mimic must be added. Lysate based expression systems have since been commercialised and used in combination with lipid-nanodiscs for functional protein and folding studies. More recently, a minimal system was developed that only contains translation components necessary for protein synthesis. This was termed PURE (protein synthesis using recombinant elements) [16]. These systems can produce large complex proteins for *in vitro* study when membrane mimics and coding DNA or mRNAs are added to the mix, and are open to a great deal of optimisation as certain components can be added or removed, for example allowing for the introduction of fluorescent labels or cross-linking reagents through unnatural-amino acid (UAA) technology. PURE is also a much cleaner approach than cytosolic extracts, facilitating the isolation and purification of affinity tagged proteins for biochemical analysis of only the protein of interest.

MP insertion has been probed using truncated or stalled RNC complexes which allow the early environmental effect on the nascent chain (NC) to be probed **(figure 1b)**. RNCs are produced by incorporating a peptide stalling sequence into a coding DNA sequence. These stalling sequences are used in nature to regulate NC translation allowing structure segments to fold upstream of the sequence [17]. The *E. coli* wild type secretion monitor (SecM) sequence, which once stalled can alter mRNA secondary structure due to ribosome binding and in turn regulate the binding of SecA to, and thus translocation of the N-terminal NC to the periplasm. The stalling occurs in the final codon of the SecM sequence, where the resulting translated proline alters the geometry of the peptidyl-transferase centre (PTC) of the ribosome [17, 18] and halts translation until nascent chain folding produces a significant

force to release the stalled sequence [19], making SecM a useful tool for RNC study [20]. The stall provided by SecM has also been enhanced based on stronger stalling sequences in other bacteria to supporting a new tool for biochemical and structural characterisation of RNCs where NC stalling release is not desired [11].

2.2. A membrane mimic

A membrane mimic with particular properties is also required for the correct folding of MPs **(figure 1b)**. Each lipid molecule has a complex chemistry, with a variable number carbon chain tail length, of which the two can be asymmetric or have different degrees of saturation. Equally, the lipid head groups provide an added physicochemical specificity. Charged head groups and non-lamellar forming lipids can affect protein topology, folding and function [3, 21-23]. In addition, amphipathic detergent molecules, often used to solubilise MPs out of the native membranes at a concentration above their critical micelle concentration (CMC) are used to encapsulate MPs by forming a micelle with their hydrophobic hydrocarbon tails facing the protein, and hydrophilic head groups facing outwards towards the aqueous buffer environment. Detergents are often the first step in MP preparation and purification and are used as a vehicle for reconstitution into synthetic lipid mixes, and in the reconstitution process itself [24].

Most studies on aspects of co-translational folding have used extracts of inner membrane vesicles (IMVs), [25-27] harvested from cells, or microsomes [28, 29], isolated from dog pancreas. More recently, nanodiscs have been employed, where lipid bilayer discs are surrounded by a membrane scaffold protein (MSP). Nanodiscs can be used with cell-free systems to produce protein that is amenable to biochemical, biophysical and structural analysis [30-32], although the MSP limits some spectroscopic analyses. Naturally, the native membrane of an organism provides the optimal environment for MP folding and function. Novel co-polymer based chemicals styrene/maleic acid (SMA) and diisobutylene/maleic acid (DIBMA) (amongst others) have been developed to purify MPs directly from native membranes without detergent. These polymers act as a 'cookie cutter' to harvest the MP and surrounding lipids and any associated proteins, making these native-nanodiscs perfect for the study of MP dynamics or folding when coupled with stalled ribosome nascent chain (RNC) complexes. These native-nanodiscs, from production through to use have been reviewed extensively elsewhere [33].

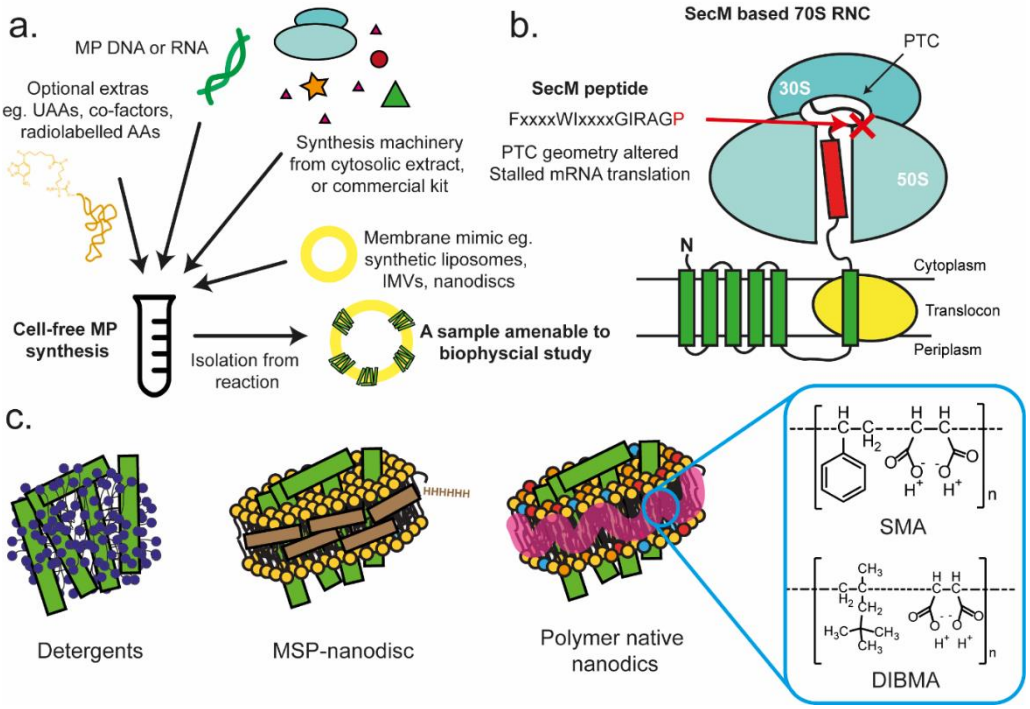

Figure 1: Summary of the necessary tools for co-translational study of MPs

a) Classic in vitro *cell-free synthesis reaction scheme. Synthesis machinery from either cytosolic extracts of an organism, or recombinant components (PURE), containing transcription and translation components (ribosomes; amino acids, tRNA and synthetases; polymerases; initiation, elongation and release factors; energy replenishment systems), are mixed with DNA or RNA of desired membrane protein, and a membrane mimic of choice to produce a sample amenable to biophysical characterisation. Additional UAAs, co-factors specific to protein function, or radiolabelled tags/amino acids can be added to the reaction mixture to label the protein of choice. The sample can be isolated from the reaction and studied, or studied using novel biophysical techniques as the reaction proceeds (see section 4). b) Simplified schematic of a SecM induced stalled 70S ribosome to produce RNCs. The minimal SecM peptide stalls its mRNA translation once the final codon for proline enters the PTC. The geometry of the PTC is altered and peptide bond formation cannot occur. The expressed, truncated nascent chain still inserts into the membrane* in vivo, *most likely using the translocon.* In vitro *it is possible for proteins to insert directly into lipids using a minimal system without the translocon present. The effect of the ribosome, and the length of produced nascent chain provide insight into co-translational insertion and folding. This latter bottom up approach is amenable to lipid composition alteration which can increase spontaneous insertion of MPs [23]. c) A selection of membrane mimics, used in protein used in co-translational protein folding. Detergents, proteinaceous MSP-nanodiscs, which are often poly-his tagged for affinity purification or tethering to a surface, and novel polymer based chemicals (SMA and DIBMA) which can purify MPs with their native membrane intact and without the need for initial detergent purification.*

3. Biochemical methods to study co-translational folding and topology

Early biochemical techniques to ascertain a time-courses for the order of polytopic TM helix insertion *in vivo*, used single-cysteine mutants amenable to labelling and were observed by a phosphorescent gel shift. Membrane impermeable 4-Acetamido-4'maleimidylstilbene-2,2'-disulfonic acid disodium salt (AMS) was used alongside ^{35}S-Methionine pulse-chase labelling to determine the insertion rates of radiolabelled bacterioopsin (bO) (bacteriorhodopsin without retinal co-factor), a well-studied 7 TM α-helical archaic proton pump of *Halobacterium salinarum* with each TM labelled A-G **(figure 2)**. A cysteine located near the N-terminus showed a bilayer translocation rate faster than the translation of full length protein, as determined by AMS derivatisation at certain time points. This suggests that the insertion of the first TM helix must occur co-translationally [34] and acts as a folding 'nucleus' for the rest of the protein. This was later advanced, and each of the extracellular loops in bO were also labelled and pulse-chased. The BC and FG loops of bO are translocated in sequence order after the first helix [34, 35], but the FG loop, unlike the BC, completes its translocation after full length protein elongation is complete, and therefore must cross the membrane post-translationally. These results together suggest a sequential order of co-translational TM insertion for bO in *H. salinarum*. Direct structure formation however, cannot be characterised using this approach.

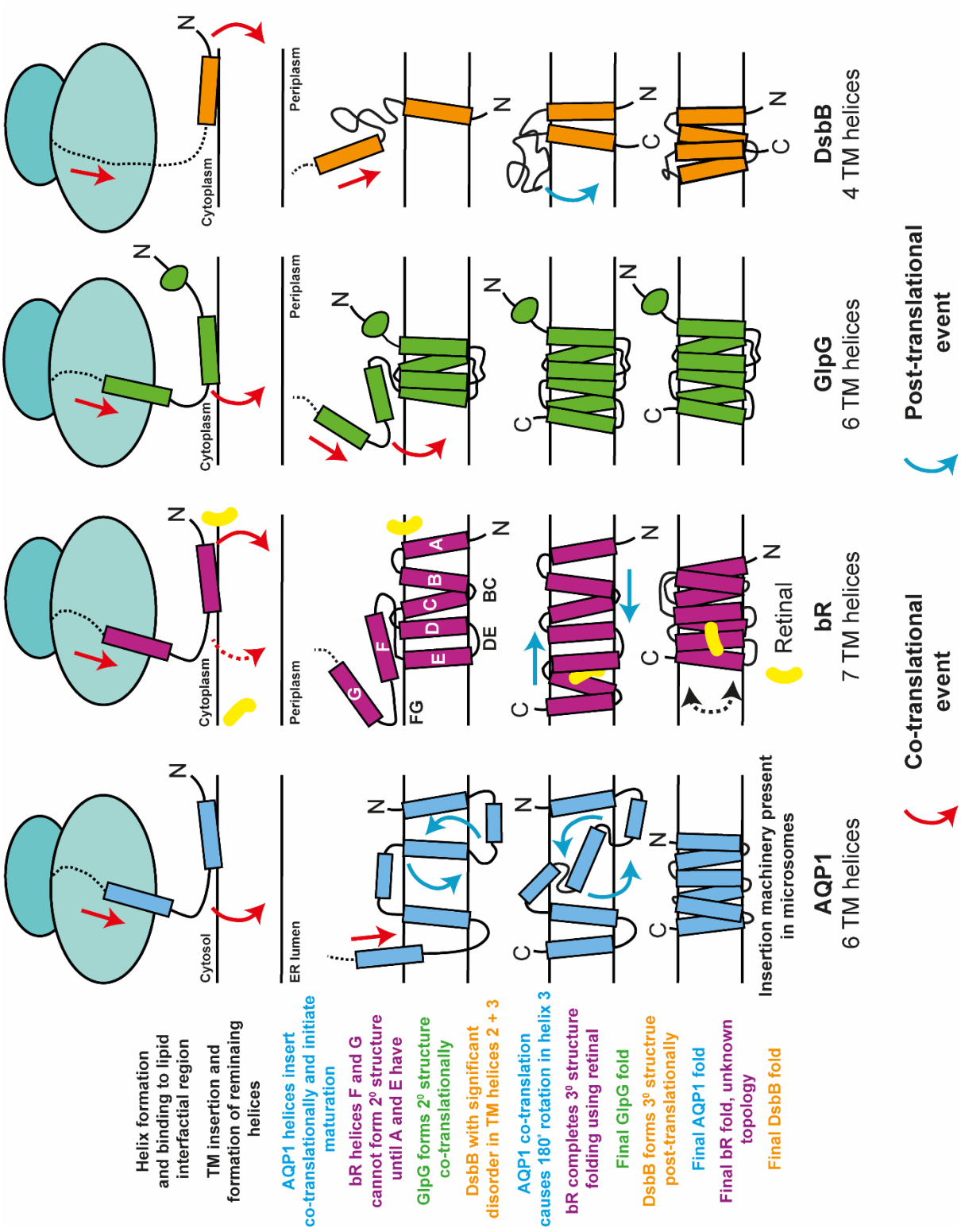

Figure 2: Possible models for co-translational folding in the absence of translocon from SEIRAS

Summary of possible models for protein folding and insertion using SEIRAS data for bR, and SEIRAS with crystal structure and mPEX predictions for GlpG and DsbB adapted from [4]. AQP1 comparison from protease reporter work in rough-ER microsomes [8]. Topology for bR in this cell-free system is unknown, the topologies of GlpG and DsbB depicted were confirmed by AMS labelling in liposomes

[4]. For each protein, helices are co-translationally (red arrows) produced and bind to the membrane interfacial region. AQP1 undergoes a co-translational maturation step where TM3 flips in the membrane 180^{o} to produce the final folded structure. Insertion machinery may aid this process in the system used. Truncated proteins were used in the AQP1 study, this may have given the NC more time to rearrange their folding pathway. bR forms helices A-E co-translationally but F and G cannot fold until A and E are correctly inserted (loops labelled for section 3). A post-translational (blue arrows) event then forms the final 3D structure of bR using retinal included in the reaction. GlpG forms secondary and tertiary structure co-translationally. DsbB Regions corresponding to TMs 2 and 3 in DsbB are not predicted to insert suggesting that the unordered regions in SEIRA correspond to TMs 2 and 3, and they insert only insert post-translationally after TMs 1 and 4. Most likely, helices 2 and 3 favour a disordered state at the DMPC interface rather than partitioning through the bilayer in the absence of any translocon machinery.

Co-translational folding of eukaryotic polytopic proteins has been studied using protease digestion to assess proteins regions outside the membrane in the cytosol. Truncated polypeptides of the 6 TM human Aquaporin-1 (AQP1) water channel [8] were fused into protease sensitive reporter constructs at TM helix-connecting loops via epitopes, which allowed determination of topology for the truncated proteins at the endoplasmic reticulum in canine pancreas rough microsomal membranes with rabbit-reticulocyte lysate (RRL). The location of protease reporter in the cytosol or ER lumen indicated the orientation of TM helices. It was determined that 4 of 6 helices initially transverse the membrane, with helices 2 and 4 binding the lipid interfacial region **(figure 2)**. A late stage maturation process appeared to flip TM helix 3 from an N-out orientation 180^{o} through the membrane to N-in orientation, in turn pulling helices 2 and 4 into the bilayer. This is a co-translational process and rotation of TM3 increases as relative C-terminal helices are released from the ribosome [8]. Other AQP1 TM segments (omitted in the diagram for clarity) appear to transiently insert into the bilayer, these regions may remain in the translocon before integration adding yet another variable to the folding pathway. The pathway for AQP1 is much more complex than the sequential insertion studies of bO. However, truncated protein chains were used to ascertain topology in AQP1, which may have given the NC extra time to alter its folded conformation before measurements have taken place. In addition, bO studies were time resolved and therefore offer a more direct measurement of co-translational folding and topology that epitope labelling can. Yet, neither study can assess structure formation during co-translational insertion nor actual folding of the protein. In other work, lipid composition has been shown to influence topology with altered *E. coli* lipids [36]. Notably lack of the main lipid phosphatidylethanolamine (PE), causes an inversion of a 6 TM domain of the 12 TM lactose permease in the membrane, which can be rescued on addition of PE.

Epitope protease assays have also been used to show that the helices themselves contain topological information in the form of signal-anchoring sequences, which open the translocon allowing translocation of loops into the ER lumen and result in an N-out helix orientation, and stop-transfer sequences, which close the translocon keeping the connecting loops on the cytosolic side of the ER membrane [37]. The strength of stop-transfer activity, as a result of hydrophilic/phobic residues in helices, and loop length, can vary resulting in extensive topology and diversity among a range of proteins. The 6 TM helix AQP4 protein regulates the co-translational folding of its final structure in this way unlike AQP1 [9], where there is no co-translational helix rotation.

As NCs are elongated, the environments of the ribosome exit tunnel and the cytosol influence folding. By incorporating fluorescent or photocrosslinking labels into the nascent chain, changes in the peptide conformation as it is produced can be monitored. UAA technology uses modified aminoacyl-tRNAs to incorporate a non-natural amino acid directly into the nascent chain during translation [38]. FRET can occur between a donor and acceptor dye when in close proximity to produce an observable fluorescence, when fluorophore based UAAs are used in translation. The efficiently of this energy transfer is proportional to the distance between the two fluorophores. This can give spatial information between

nascent chain positions, and when photo-crosslinking agents are introduced, interactions between the unnatural amino acids and translocon components are observed [38, 39].

Single membrane-spanning TM peptides appear to fold near the ribosome PTC to produce compact α-helices. This is shown by a high FRET efficiency. The compact nature is lost as the NC is released into the exit-tunnel, characterised by a reduced FRET efficiency. The final membrane inserted TM shows a similar FRET efficiency to when positioned near the PTC. Together, this shows that the ribosome itself can play a role in the initial folding strategy for TM spanning proteins. This pathway was proven to be orchestrated by specific ordered ribosome/TM interactions using photo-crosslinking reagents. The TM segment was shown to crosslink with three eukaryotic ribosomal proteins [39], whereas non-TM helix sequences crosslinked to only one. This suggests that the ribosome contains specific sequence selectors for TM helices, with the first acting as an initial sensor for TM segments, to provide a hydrophobic nucleation point for folding [39], indicating that folding pathways are dependent on sequence, and protein identity. These interactions further show that specific sequences in the TM-helix primary structure can control the entry of the helix into the translocon much like the epitope labelling work, however UAAs have only been used to study single-membrane spanning proteins with the overall picture of insertion of polytopic proteins still to be uncovered. Certain fluorescence dyes coupled with UAA technology can also give insightful information into its hydrophobic microenvironment due to a red or blue shift in its emission maxima. This was nicely used to uncover minor structural alterations in the TIM23 mitochondrial channel when subject to a proton motive force (PMF) using wheat-germ extract, an alternative cell-free expression system sometimes used for eukaryotic protein expression [40].

The SecM arrest-peptide (AP) has been used as an *in vivo* force-sensor by the von Heijne group, giving insight into forces acting on the co-translationally inserting and folding segments of the NC [41, 42]. The force of NC release is related to a measurement of the fraction of truncated protein per fully translated protein (f_{FL}), and plotted against helix number, and a force-profile for a particular protein is determined. This methodology seeks to describe the force (or f_{FL}) acting on a hydrophobic segment at varying distance from the AP, most likely describing interactions arising from NC insertion into the translocation machinery and TM helix partitioning into the membrane.

It was hypothesised that the detected force experienced by hydrophobic domains increases when the upstream TM helices fold and partition into the membrane, possibly signifying interactions between helices in the co-translational folding process [43]. Polytopic membrane proteins; CaiT, NhaA, EmrD, BtuC and GlpT, were cloned into the AP construct [44]. The apparent free-energies of protein insertion (ΔG_{app}) were predicted using ΔG-predictor [45] and of each candidate, one TM with a positive predicted ΔG_{app} was selected suggesting unfavourable helix insertion. The interactions between the helix and the upstream N-terminal helices were quantified by an f_{FL} value, and show that an increased hydrophobicity subsequently increased insertion, as did the presence of previously synthesised upstream helices. The shorter the loop length between the helix and the prior helices also increased f_{FL} signifying the reason for the short loop length in polytopic membrane proteins. The exact identity each of these interactions themselves however cannot be determined, nor can they give us a time resolved sequence of events of folding and tertiary structure formation on the growing polypeptide.

Furthermore, cryo-electron microscopy has been used to visualise static interactions between 7 TM proteorhodopsin (PR) [46], and the SecY translocon machinery in an RNC stalled state. A tryptophan dependent TnaC stalling peptide was used to generate SecY-bound RNCs with two complete TM helices and a hydrophilic region, necessary for stalling and producing tight interactions with SecY. These RNCs were co-purified in detergent with SecYEG and a RNC-SecY structure was obtained to 7.3 Å. The structure supports the idea that the TM helices may exit the SecY channel via the lateral gate for partition into the membrane as TMs 1 and 2 are directly in front of the lateral gate. It was proposed that YidC also positions itself outside of the lateral gate (although not detected in this study), suggesting

further chaperone regulation for TM partitioning, along with SecA, which cannot bind to the ribosome at the same time as SecY. Interactions between the positively charged cytoplasmic loop of the first two helices with a ribosomal rRNA helix were also observed. This suggests a role for the ribosome in retaining charged loops on the cytoplasmic side of the bilayer during TM integration into the membranes in this system.

Membrane protein co-translational folding studies currently lag behind that of soluble folding studies. Nascent chain studies of the latter, soluble proteins include HemK folding pathways by the Rodnina lab, using *in vitro* FRET based assays between fluorophores at different positions along the nascent chain [47]. The changes in FRET were time-resolved and suggested 4 independent intermediates for the co-translational folding of HemK which begin in the ribosome exit tunnel, and are only limited by translation rate. Translation rate and its effect on folding has been further explored harnessing rare codon technology [48]. Furthermore, much progress has been made in the structural dynamics of trapped nascent chains by the Christodoulou group using various NMR techniques with SecM mediated stalling [49, 50]. The globular FLN5 protein domain sampled transient interactions with the ribosomal RNA and the proteins surrounding the exit site.

4. Time-resolving nascent chain structure formation during co-translational folding

A recent approach to directly observe temporal formation of NC structure, during co-translational folding of MPs into membranes is surface enhanced IR spectroscopy (SEIRAS). The SEIRA setup **(figure 3)** allows for IR spectra to be in the field of IR enhancement, resulting from plasmonic resonance at a thin gold surface deposited on a silicon prism. This gives a 10-100-fold increase in sensitivity compared to conventional IR spectroscopy [51]. There is a 10 nm enhancement region, which allows spectra to be obtained only of uniformly oriented lipid-nanodiscs and protein within the nanodisc, but not the translating ribosomes nor anything outside the membrane. The IR amide I (C=O stretching vibration) and II band (C-N stretching and N-H bending) intensities give fingerprints of protein structure, and particularly the position and width of the amide I band indicate helix formation packing during co-translational insertion.

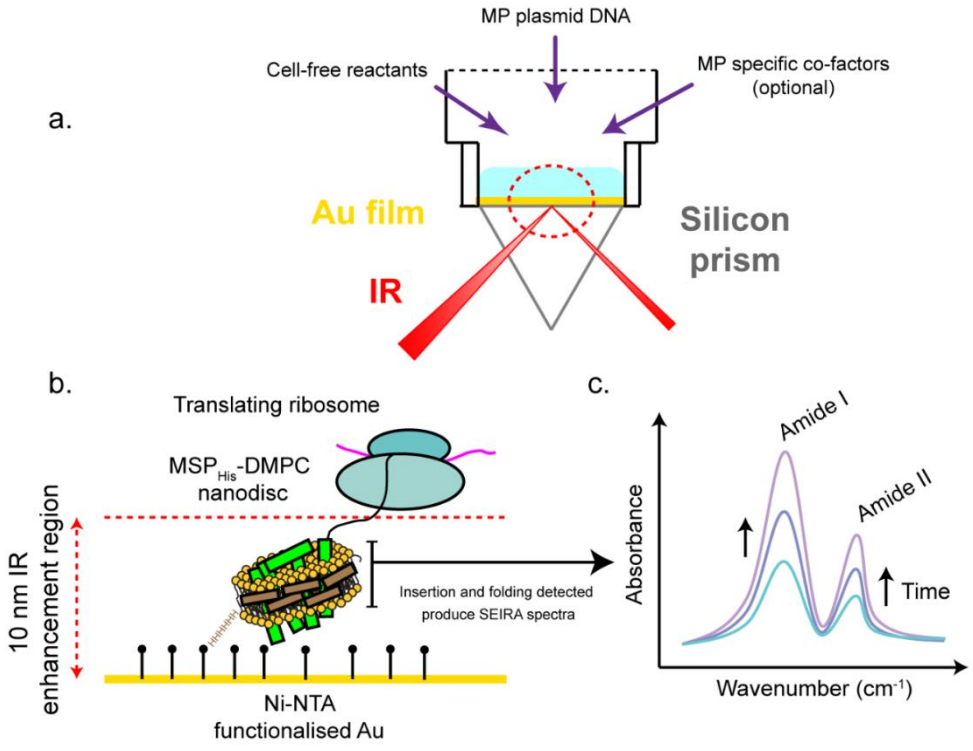

Figure 3: SEIRAS set-up and methodology

a) Reaction chamber and contents. The chamber sits on silicon prism with a thin-layer of gold. A 10 nm enhancement area above the gold surface results from plasmon resonance, due a reflection of the IR beam in the gold-prism interface. Plasmid DNA of the MP of interest is added, along with the cell-free kit containing ribosomes, T7 polymerase and 'feed' mixture with amino acids and energy replenishment components. Here, any co-factors required for MP folding (such as retinal for bR) are added. b) The IR beam illuminates a region up to 10 nm where a uniform layer of MSP-DMPC poly-histidine tethered nanodiscs sit on the Ni-NTA functionalised gold. The ribosomes translate the plasmid DNA producing NCs, which spontaneously insert into the nanodisc allowing in situ *measurements of co-translational folding. c) Schematic SEIRA spectra output showing amide I and amide II peak intensity, which increase over time, and yielding kinetic information on folding.*

The first MP co-translational study was carried out on bacteriorhodopsin (bR), using a commercial cell-free kit with bR DNA, retinal cofactor, and 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) based MSP-nanodiscs. Both amide I and II bands increased as a function of time after the reaction started until a significant shift in peaks suggested that membrane insertion had occurred. In the late stages of folding, tertiary structure formation was observed with a sharpening of the amide I band, characteristic of α-helical bundling [52]. No structure formation was observed when retinal was absent.

More recently, cell-free expression of the *E. coli* membrane proteins rhomboid protease GlpG and disulphide reductase DsbB were followed by SEIRAS. As with bO, no translocon was present and thus the observed folding was spontaneous and efficient, unassisted by translation apparatus [4]. For GlpG, the earliest detected IR bands corresponded to α-helices and β-structure, and disordered structure. The latter then decreased as more α-helical structure was produced. α-helical bundling between TM helices occurred [4, 52] as GlpG folds into its polytopic structure. These structural changes all occurred within the time required to make the full length protein, strongly indicating that all folding occurred co-translationally.

DsbB co-translational folding was slower with more disordered structure initially, and helix formation and packing occurring both co- and post-translationally [4]. The observed structure formation during folding of both GlpG and DsbB was consistent with hydropathy plots; all 6 helices of GlpG are predicted to be stable by hydrophobicity scales for insertion into the lipid headgroups as well as via the translocon [53]. In contrast, TMs 2 and 3 of the 4 TM DsbB are predicted to be unstable, and thus could initially form disordered structure and require TMs 1 and 4 already in the membrane for the remaining helices to insert and fold to give the final protein structure summarised in **(figure 2)**. DsbB however, contains a large unstructured periplasmic loop, which does not cross the membrane in this reductionist system. This results in an inverted topology as shown by AMS labelling in liposomes. *In vivo*, the translocon is likely to translocate the periplasmic loop across the membrane. SEIRAS itself, without labelling of the protein, cannot identify which regions of the protein are folding.

Each protein tested using SERIAS so far has inserted into the bilayer in the absence of any translocation machinery suggesting, and confirming previous work [54-57] that the lipids themselves possess the necessary mechanical energy to modulate lipid binding independent of the translocon. The co-translational folding of GlpG for example was enhanced by certain lipids, including PE which alters the mechanical properties and phosphatidylglycerol (PG) that introduces negative headgroup charge. There is a degree of folded and misfolded structure for GlpG, DsbB and bR (**figure 2)** in their early stages, and it may be that addition of a translocon would reduce the amount of misfolded protein.

6. Conclusions

There are currently very few studies on the structure formation of α-helical membrane proteins from a co-translational perspective. However, our ever expanding and refined arsenal of tools to study these systems, coupled with both biochemical and biophysical methods, is starting to paint a creditable picture

of the folding pathways of a handful of polytopic membrane proteins. Bacteriorhodopsin folding has been extensively studied through classic denaturant refolding experiments and biochemical assays to determine the order of TM insertion, and subsequently by time-resolved SEIRAS. This latter approach has considerable potential to reveal the temporal formation of structure during co-translational insertion and folding, particularly when combined with other approaches such as FRET, protease digestion and RNCs. The details of secondary structure formation for GlpG and DsbB has been observed and refined with input from predicted hydropathy plots. Additionally, biochemical principles of using *in vivo* force-mediated sensors can give insight into insertion, whilst cryoEM is beginning to deliver high resolution structural information of nascent chains and their local interactions with the insertion apparatus.

Acknowledgements

We thank Dr Nicola Harris, Dr Heather Findlay and Dr Eamonn Reading for proofreading the manuscript.

We thank the ERC and King's College, London for funding (Advanced Grant 294342 to PJB and PhD studentship to GP).

References

[1] V.A. Gold, F. Duong, I. Collinson, Structure and function of the bacterial Sec translocon, Mol Membr Biol, 24 (2007) 387-394, DOI: 10.1080/09687680701416570
[2] J. Komar, S. Alvira, R.J. Schulze, R. Martin, A.N.J.A. Lycklama, S.C. Lee, T.R. Dafforn, G. Deckers-Hebestreit, I. Berger, C. Schaffitzel, I. Collinson, Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG-SecDF-YajC-YidC, Biochem J, 473 (2016) 3341-3354, DOI: 10.1042/BCJ20160545
[3] W. Dowhan, M. Bogdanov, Lipid-protein interactions as determinants of membrane protein structure and function, Biochem Soc Trans, 39 (2011) 767-774, DOI: 10.1042/BST0390767
[4] N.J. Harris, E. Reading, K. Ataka, L. Grzegorzewski, K. Charalambous, X. Liu, R. Schlesinger, J. Heberle, P.J. Booth, Structure formation during translocon-unassisted co-translational membrane protein folding, Sci Rep-Uk, 7 (2017), DOI: ARTN 8021 10.1038/s41598-017-08522-9
[5] T.A. Rapoport, Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes, Nature, 450 (2007) 663-669, DOI: 10.1038/nature06384
[6] M. Hermansson, G. von Heijne, Inter-helical hydrogen bond formation during membrane protein integration into the ER membrane, Journal of Molecular Biology, 334 (2003) 803-809, 10.1016/j.jmb.2003.10.019
[7] D. Pitonzo, W.R. Skach, Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon, Biochim Biophys Acta, 1758 (2006) 976-988, DOI: 10.1016/j.bbamem.2006.04.021
[8] Y. Lu, I.R. Turnbull, A. Bragin, K. Carveth, A.S. Verkman, W.R. Skach, Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum, Mol Biol Cell, 11 (2000) 2973-2985, DOI: 10.1091/mbc.11.9.2973
[9] W. Foster, A. Helm, I. Turnbull, H. Gulati, B.X. Yang, A.S. Verkman, W.R. Skach, Identification of sequence determinants that direct different intracellular folding pathways for aquaporin-1 and aquaporin-4, Journal of Biological Chemistry, 275 (2000) 34157-34165, DOI: 10.1074/jbc.M000165200
[10] D.V. Fedyukina, S. Cavagnero, Protein Folding at the Exit Tunnel, Annual Review of Biophysics, Vol 40, 40 (2011) 337-359, DOI: 10.1146/annurev-biophys-042910-155338
[11] F. Cymer, G. von Heijne, S.H. White, Mechanisms of integral membrane protein insertion and folding, J Mol Biol, 427 (2015) 999-1022, DOI: 10.1016/j.jmb.2014.09.014
[12] B. Martoglio, M.W. Hofmann, J. Brunner, B. Dobberstein, The Protein-Conducting Channel in the Membrane of the Endoplasmic-Reticulum Is Open Laterally toward the Lipid Bilayer, Cell, 81 (1995) 207-214, DOI: 10.1016/0092-8674(95)90330-5

[13] R.E. Dalbey, A. Kuhn, L. Zhu, D. Kiefer, The membrane insertase YidC, Biochim Biophys Acta, 1843 (2014) 1489-1496, DOI: 10.1016/j.bbamcr.2013.12.022
[14] J. Serek, G. Bauer-Manz, G. Struhalla, L. van den Berg, D. Kiefer, R. Dalbey, A. Kuhn, Escherichia coli YidC is a membrane insertase for Sec-independent proteins, EMBO J, 23 (2004) 294-301, DOI: 10.1038/sj.emboj.7600063
[15] P. Wang, R.E. Dalbey, Inserting membrane proteins: the YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts, Biochim Biophys Acta, 1808 (2011) 866-875, 10.1016/j.bbamem.2010.08.014
[16] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, Cell-free translation reconstituted with purified components, Nat Biotechnol, 19 (2001) 751-755, DOI: 10.1038/90802
[17] K. Ito, S. Chiba, K. Pogliano, Divergent stalling sequences sense and control cellular physiology, Biochem Biophys Res Commun, 393 (2010) 1-5, DOI: 10.1016/j.bbrc.2010.01.073
[18] S. Bhushan, T. Hoffmann, B. Seidelt, J. Frauenfeld, T. Mielke, O. Berninghausen, D.N. Wilson, R. Beckmann, SecM-Stalled Ribosomes Adopt an Altered Geometry at the Peptidyl Transferase Center, Plos Biol, 9 (2011), DOI: ARTN e1000581 10.1371/journal.pbio.1000581
[19] N. Ismail, R. Hedman, M. Linden, G. von Heijne, Charge-driven dynamics of nascent-chain movement through the SecYEG translocon, Nat Struct Mol Biol, 22 (2015) 145-149, DOI: 10.1038/nsmb.2940
[20] M.S. Evans, K.G. Ugrinov, M.A. Frese, P.L. Clark, Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro, Nat Methods, 2 (2005) 757-762, 10.1038/nmeth790
[21] P.J. Booth, P. Curnow, Folding scene investigation: membrane proteins, Curr Opin Struct Biol, 19 (2009) 8-13, DOI: 10.1016/j.sbi.2008.12.005
[22] H.E. Findlay, P.J. Booth, The folding, stability and function of lactose permease differ in their dependence on bilayer lipid composition, Sci Rep-Uk, 7 (2017), DOI: ARTN 13056 10.1038/s41598-017-13290-7
[23] N.J. Harris, K. Charalambous, H.E. Findlay, P.J. Booth, Lipids modulate the insertion and folding of the nascent chains of alpha helical membrane proteins, Biochem Soc T, 46 (2018) 1355-1366, DOI: 10.1042/Bst20170424
[24] A.M. Seddon, P. Curnow, P.J. Booth, Membrane proteins, lipids and detergents: not just a soap opera, Bba-Biomembranes, 1666 (2004) 105-117, DOI: 10.1016/j.bbamem.2004.04.011
[25] Y. Kuruma, K. Nishiyama, Y. Shimizu, M. Muller, T. Ueda, Development of a minimal cell-free translation system for the synthesis of presecretory and integral membrane proteins, Biotechnol Prog, 21 (2005) 1243-1251, DOI: 10.1021/bp049553u
[26] J.J. Wuu, J.R. Swartz, High yield cell-free production of integral membrane proteins without refolding or detergents, Bba-Biomembranes, 1778 (2008) 1237-1250, DOI: 10.1016/j.bbamem.2008.01.023
[27] K. Nishiyama, M. Maeda, M. Abe, T. Kanamori, K. Shimamoto, S. Kusumoto, T. Ueda, H. Tokuda, A novel complete reconstitution system for membrane integration of the simplest membrane protein, Biochem Bioph Res Co, 394 (2010) 733-736, DOI: 10.1016/j.bbrc.2010.03.061
[28] R.B. Quast, O. Kortt, J. Henkel, S.K. Dondapati, D.A. Wustenhagen, M. Stech, S. Kubick, Automated production of functional membrane proteins using eukaryotic cell-free translation systems, J Biotechnol, 203 (2015) 45-53, DOI: 10.1016/j.jbiotec.2015.03.015
[29] R.B. Quast, A. Sonnabend, M. Stech, D.A. Wustenhagen, S. Kubick, High-yield cell-free synthesis of human EGFR by IRES-mediated protein translation in a continuous exchange cell-free reaction format, Sci Rep-Uk, 6 (2016), DOI: ARTN 30399 10.1038/srep30399
[30] T.H. Bayburt, Y.V. Grinkova, S.G. Sligar, Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins, Nano Lett, 2 (2002) 853-856, DOI: 10.1021/nl025623k

[31] T.H. Bayburt, S.G. Sligar, Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks, P Natl Acad Sci USA, 99 (2002) 6725-6730, DOI: 10.1073/pnas.062565599
[32] I.G. Denisov, Y.V. Grinkova, A.A. Lazarides, S.G. Sligar, Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size, J Am Chem Soc, 126 (2004) 3477-3487, DOI: 10.1021/ja0393574
[33] K.S. Simon, N.L. Pollock, S.C. Lee, Membrane protein nanoparticles: the shape of things to come, Biochem Soc Trans, 46 (2018) 1495-1504, DOI: 10.1042/BST20180139
[34] H. Dale, M.P. Krebs, Membrane insertion kinetics of a protein domain in vivo. The bacterioopsin n terminus inserts co-translationally, J Biol Chem, 274 (1999) 22693-22698, DOI: 10.1074/jbc.274.32.22693
[35] H. Dale, C.M. Angevine, M.P. Krebs, Ordered membrane insertion of an archaeal opsin in vivo, Proc Natl Acad Sci U S A, 97 (2000) 7847-7852, DOI: 10.1073/pnas.140216497
[36] M. Bogdanov, P.N. Heacock, W. Dowhan, A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition, EMBO J, 21 (2002) 2107-2116, DOI: 10.1093/emboj/21.9.2107
[37] K. Moss, A. Helm, Y. Lu, A. Bragin, W.R. Skach, Coupled translocation events generate topological heterogeneity at the endoplasmic reticulum membrane, Molecular Biology of the Cell, 9 (1998) 2681-2697, DOI: 10.1091/mbc.9.9.2681
[38] A.E. Johnson, The co-translational folding and interactions of nascent protein chains: a new approach using fluorescence resonance energy transfer, Febs Lett, 579 (2005) 916-920, DOI: 10.1016/j.febslet.2004.11.046
[39] C.A. Woolhead, P.J. McCormick, A.E. Johnson, Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins, Cell, 116 (2004) 725-736, DOI: 10.1016/S0092-8674(04)00169-2
[40] N.N. Alder, R.E. Jensen, A.E. Johnson, Fluorescence mapping of mitochondrial TIM23 complex reveals a water-facing, substrate-interacting helix surface, Cell, 134 (2008) 439-450, DOI: 10.1016/j.cell.2008.06.007
[41] M.J.M. Niesen, A. Muller-Lucks, R. Hedman, G. von Heijne, T.F. Miller, 3rd, Forces on Nascent Polypeptides during Membrane Insertion and Translocation via the Sec Translocon, Biophys J, 115 (2018) 1885-1894, DOI: 10.1016/j.bpj.2018.10.002
[42] D.H. Goldman, C.M. Kaiser, A. Milin, M. Righini, I. Tinoco, Jr., C. Bustamante, Ribosome. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo, Science, 348 (2015) 457-460, DOI: 10.1126/science.1261909
[43] N. Ismail, R. Hedman, N. Schiller, G. von Heijne, A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration, Nat Struct Mol Biol, 19 (2012) 1018-U1068, DOI: 10.1038/nsmb.2376
[44] F. Cymer, G. von Heijne, Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements, Proc Natl Acad Sci U S A, 110 (2013) 14640-14645, DOI: 10.1073/pnas.1306787110
[45] T. Hessa, N.M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S.H. White, G. von Heijne, Molecular code for transmembrane-helix recognition by the Sec61 translocon, Nature, 450 (2007) 1026-1030, DOI: 10.1038/nature06387
[46] L. Bischoff, S. Wickles, O. Berninghausen, E.O. van der Sluis, R. Beckmann, Visualization of a polytopic membrane protein during SecY-mediated membrane insertion, Nat Commun, 5 (2014), DOI: ARTN 4103 10.1038/ncomms5103
[47] W. Holtkamp, G. Kokic, M. Jager, J. Mittelstaet, A.A. Komar, M.V. Rodnina, Cotranslational protein folding on the ribosome monitored in real time, Science, 350 (2015) 1104-1107, 10.1126/science.aad0344

[48] F. Buhr, S. Jha, M. Thommen, J. Mittelstaet, F. Kutz, H. Schwalbe, M.V. Rodnina, A.A. Komar, Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations, Mol Cell, 61 (2016) 341-351, 10.1016/j.molcel.2016.01.008
[49] A.M.E. Cassaignau, H.M.M. Launay, M.E. Karyadi, X.L. Wang, C.A. Waudby, A. Deckert, A.L. Robertson, J. Christodoulou, L.D. Cabrita, A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy, Nat Protoc, 11 (2016) 1492-1507, DOI: 10.1038/nprot.2016.101
[50] L.D. Cabrita, A.M.E. Cassaignau, H.M.M. Launay, C.A. Waudby, T. Wlodarski, C. Camilloni, M.E. Karyadi, A.L. Robertson, X.L. Wang, A.S. Wentink, L.S. Goodsell, C.A. Woolhead, M. Vendruscolo, C.M. Dobson, J. Christodoulou, A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding, Nat Struct Mol Biol, 23 (2016) 278-285, 10.1038/nsmb.3182
[51] K. Ataka, S.T. Stripp, J. Heberle, Surface-enhanced infrared absorption spectroscopy (SEIRAS) to probe monolayers of membrane proteins, Biochim Biophys Acta, 1828 (2013) 2283-2293, DOI: 10.1016/j.bbamem.2013.04.026
[52] A. Baumann, S. Kerruth, J. Fitter, G. Buldt, J. Heberle, R. Schlesinger, K. Ataka, In-Situ Observation of Membrane Protein Folding during Cell-Free Expression, PLoS One, 11 (2016) e0151051, DOI: 10.1371/journal.pone.0151051
[53] C. Snider, S. Jayasinghe, K. Hristova, S.H. White, MPEx: A tool for exploring membrane proteins, Protein Sci, 18 (2009) 2624-2628, DOI: 10.1002/pro.256
[54] P.J. Booth, R.H. Templer, W. Meijberg, S.J. Allen, A.R. Curran, M. Lorch, In vitro studies of membrane protein folding, Crit Rev Biochem Mol Biol, 36 (2001) 501-603, 10.1080/20014091074246
[55] P. Curnow, P.J. Booth, Combined kinetic and thermodynamic analysis of alpha-helical membrane protein unfolding, Proc Natl Acad Sci U S A, 104 (2007) 18970-18975, DOI: 10.1073/pnas.0705067104
[56] H. Lu, P.J. Booth, The final stages of folding of the membrane protein bacteriorhodopsin occur by kinetically indistinguishable parallel folding paths that are mediated by pH, J Mol Biol, 299 (2000) 233-243, DOI: 10.1006/jmbi.2000.3735
[57] J.U. Bowie, Membrane proteins: a new method enters the fold, Proc Natl Acad Sci U S A, 101 (2004) 3995-3996, DOI: 10.1073/pnas.0400671101